Reviewing Rasch decisions

Back in the 60s, British statisticians vigorously rejected a new analytical technique developed by Danish mathematician Georg Rasch. In so doing, they condemned the UK to an analytical backwater, argues Peter Tymms.

In the 1950s, Danish mathematician Georg Rasch was presented with children's reading test data and asked to analyse it for the Danish Ministry of Social Affairs. Quite ignorant of the traditional way of conducting test analyses, Rasch went ahead and developed his own approach, his simple yet elegant solution producing some remarkable insights and marking a breakthrough in test data analysis.

Out with the old?

For much of the twentieth century, analysts had relied on an approach known as Classical Test Theory (CTT) to guide their thinking when looking at data from examinations and other kinds of assessment. But as far back as the 1920s, psychometricians had highlighted the technique's shortcomings, and worried about the problems that CTT sidestepped. An interesting perspective on this comes from Tom Bramley's recent paper on the "paired comparison methods" which revisits some of the early ideas and gives them a modern use in the examining context and a basis in Rasch theory.

Rasch's new technique provided a tool that addressed many of CTT's failings. For example, he put question difficulty and the ability of those who took the test onto the same scale – a very useful association referred to as conjoint measurement; it allows teacher and others to judge just how difficult each item is for every testee and vice versa. Rasch analysis also did away with the assumption of normal distributions, something classical analysts had drawn strength from but which had the potential to mislead both the analyst and lead to questions being added or removed from tests unnecessarily.

Rasch measurement also allowed for the difficulty of a test to be separated from the ability of the test-takers. Whereas CTT generates questions of a difficulty related to the profile of the group being tested, the new approach meant that if an assessment were given to one group of subjects and then to another group of subjects, the difficulties of the questions would remain constant.

Finally Rasch measurement creates an equal interval scale. No one would claim that the difference between grades A and B in at GCSE is equivalent to the difference between grades C and a D. It does not even seem that it would be possible to make them equal, but Rasch shows how this could be done.

These major characteristics led theorists to consider the fundamentals of 'measurement', and to dismiss the suggestion that simply adding up numbers on a test or questionnaire fulfilled the criteria.

Rasch reaction

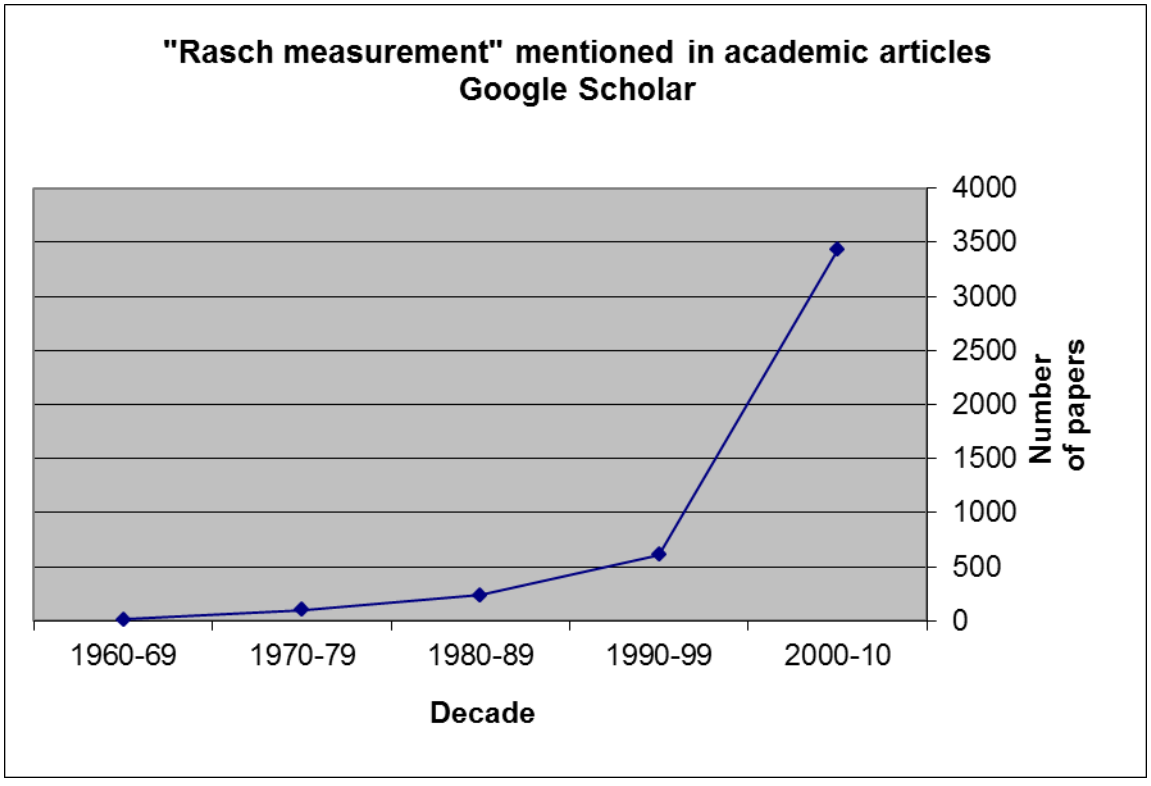
Rasch's ideas were revolutionary, and when we have revolutions we expect sparks to fly and arguments to be fought.

When Bruce Choppin brought these concepts to England in the 1960s and proposed creating question banks with known difficulties, his approach was attacked with vigour. It was claimed for example that context was so important that isolated difficulties could not be ascribed to items and that pure single construct scales were not possible amongst in education. The traditionalists won the battle, and such was the fallout that Rasch was banished from the National Foundation for Education Research where it had seen a brief flourishing.

This stark reaction to the technique 'condemned Britain to a 60 year regression', according to Mike Linacre, a leading proponent of Rasch methodology. Rasch was not mentioned in any article in the *British Educational Research Journal* from 1980 until 2010, when a paper by Panayides *et al.* recounted the history and the responses to the major criticisms of the model. Fortunately Rasch's ideas were developed and taken up by far-sighted workers elsewhere around the world, key amongst them were Mike Linacre and Geoff Masters of the University of Chicago at the time, and David Andrich of the University of Western Australia.

Rasch today

The chart below shows how often the phrase 'Rasch measurement' has been used in academic articles over five decades according to Google Scholar. The growth is dramatic, and in 2012 the phrase appeared 853 times – averaging out at just over two published articles per day!



The UK Rasch User Group, established in 2006, is part of a drive to reinvigorate a UK movement which was started in the 1960s and then essentially suppressed for decades. Rasch analysis has been carried forward by key players who have done much to resolve crucial issues and extend others, but there is still much work to be done. It has had a major impact on international country-wide tests, on computer adaptive testing and in the thinking of psychometrians but we are still at a crossroads and much is round the corner in terms of applications.

References

Andrich, D. (2004). Controversy and the Rasch Model: A Characteristic of Incompatible Paradigms? *Medical Care*, *42*(1 suppl), I-7 to I-16.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2nd edition ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates

Bramley, T. 2007 : *Paired comparison methods*, in: P. Newton, J. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.) Techniques for monitoring the comparability of examination standards. London: QCA

Linacre, J.M. (1995). *Bruce Choppin: visionary*. Retrieved 28 January 2013.

Panayides, P., Robinson, C., & Tymms, P. (2009). The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal*. Retrieved 28 January 2013.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests* (reproduced in 1980 with a foreword and afterword by Benjamin D. Wright). Chicago, IL: University of Chicago.